METHOD AND SYSTEM OF PRODUCT DEVELOPMENT PROCESS
FOR CHEMICAL COMPOSITIONS USING HIGH VOLUME MODELING

BACKGROUND OF THE INVENTION

1. Technical Field

The present invention relates generally to a method and system of product

development process for chemical compositions using high volume modeling.

2. Description of the Related Art

The use of a combinatorial approach for materials synthesis is a relatively new

area of research aimed at using rapid synthesis and screening methods to build libraries of

polymeric, inorganic or solid state materials. For example, advances in reactor

technology have empowered chemists and engineers to rapidly produce large libraries of

discrete organic molecules in the pursuit of new drug discovery, which have led to the

development of a growing branch of research called combinatorial chemistry.

Combinatorial chemistry generally refers to methods and materials for creating

collections of diverse materials or compounds--commonly known as libraries--and to

techniques and instruments for evaluating or screening libraries for desirable properties.

Presently, research in the lubricant industry involves individually forming

candidate lubricating oil compositions and then performing a macro-scale analysis of the

candidate compositions by employing a large amount of the candidate to be tested.

Additionally, the methods employed for testing each candidate composition require manual

operation. This, in turn, significantly reduces the number of compositions that can be

tested and identified as leading lubricating oil compositions.

Drawbacks associated with conventional screening procedures can be seen as follows. For example, governmental and automotive industry pressure towards reducing the phosphorous and sulfur content of lubricating oil compositions is leading to new research to identify oil compositions which can satisfy certain tests such as, for example, oxidation, wear and compatibility tests, while containing low levels of phosphorous and sulfur. For instance, United States Military Standards MIL-L-46152E and the ILSAC Standards defined by the Japanese and United States Automobile Industry Association at present require the phosphorous content of engine oils to be at or below 0.10 wt. % with future phosphorous content being proposed to even lower levels, e.g., 0.08 wt. % by January, 2004 and below 0.05 wt. % by January, 2006. Also, at present, there is no industry standard requirement for sulfur content in engine oils, but it has been proposed that the sulfur content be below 0.2 wt. % by January, 2006. Thus, it would be desirable to decrease the amount of phosphorous and sulfur in lubricating oils still further, thereby meeting future industry standard proposed phosphorous and sulfur contents in the engine oil while still retaining the oxidation or corrosion inhibiting properties and antiwear properties of the higher phosphorous and sulfur content engine oils.

Typically, once a group of lubricating oil compositions selected to address a certain requirement, for example the amount of phosphorous and sulfur, has been prepared, no additional tests related to numerous properties of the oil compositions of the selected group, are performed. As a consequence, oftentimes testing additional properties of the selected group of oil compositions, which include, for example, the properties tested during a variety of wear tests, comes as an afterthought. Yet, ensuring proper antiwear properties of lubricating oil compositions are critical to successful operation and

maintenance of mechanical systems, such as automobile engines. Laboratory lubricant analysis is time consuming and costly. Furthermore, the laboratories are typically not sufficiently automated, which leads to performing the same lengthy procedure every time a new lubricating oil composition is being tested.

Thus, present research in the lubricant industry does not allow for diverse and rapid testing of lubricating oil compositions. As such, there exists a need in the art for a more efficient, economical and systematic approach for the bench testing of lubricating oil compositions and screening of such compositions for information potentially bearing upon the actual useful properties of the compositions.

Accordingly, it would be desirable to rapidly screen a plurality of sample candidate lubricating oil compositions utilizing small amounts to automatically determine and catalog the desired lubricating properties. In this manner, a method and system of product development process for chemical compositions using high volume modeling can be achieved.

SUMMARY OF THE INVENTION

The key variables with regard to the efficacy of modeling approaches are the relative complexity of the basic product from a molecular point of view and the number of variations from the basic product, e.g., formulations. Lubricating oil additives are characterized by relatively complex products and numerous variations and formulations of the basic components. In addition, throughout the product development chain, lubricating oil additives are characterized by a series of often relatively poor correlating tests, but with the ultimate testing platform, e.g., engine tests, being very expensive. In

such a complex, noisy test environment, molecular-level modeling will be the least effective approach, although it may facilitate innovation with regard to radically different chemistries. Advanced, comprehensive statistical inferencing is guaranteed to be critical in this environment, but limited by the number of physical test points and inherent noise. High throughput experimentation (HTE) seems particularly well suited to lubricating oil additives or lubricating oil compositions containing such additives, as in combination with statistical inferencing, it can generate the large number of physical tests required to increase the depth and breadth of the product "formulation landscape". HTE and associated advanced statistical inferencing is ideally suited for lubricating oil additives and formulated lubricating oil compositions.

The present invention is guided by making improvements in the ability to make confident inferences from test results and the volume of tests on formulation variations that can be conducted. In the molecular modeling step there is an opportunity, through enhanced processes, to achieve benefits with regard to improving test inferencing confidence. In the comprehensive statistical inferencing step, this can be supplemented by application of quantitative structure activity relationship (QSAR) techniques. In the HTE step, basic HTE is applied, with bench tests at current correlation levels miniaturized and included within the HTE infrastructure. In the bench test step, miniaturized tests with better correlation than existing bench tests will be integrated into the HTE infrastructure.

The present invention describes a method and system of transforming a product development process to reduce time to bring a product to market through high throughput experimentation and advanced statistics and informatics. The present invention

transforms the product development process to a level of higher correlation with engine tests, and to develop better commercial products. This is achieved by in Silico modeling of a plurality of component molecular models using quantum mechanical (QM) software method; in Silico deriving molecular characteristics (descriptors) for each of the plurality of compiled molecular models performed by QSAR software. As used herein, the expression "in Silico" or "in Silico modeling" refers to a computational work.

The present invention further includes at least creating at least one combinatorial library database record for each of the formulations, and at least one record having a plurality of fields for storing information about compositional characteristics. The libraries include component combinations and an entire mixture of lubricating oil additives or lubricating oil compositions containing at least one base oil of lubricating viscosity and at least one lubricating oil additive. The formation and management is performed by a software program managing a relational database, such as, e.g., Oracle.

New lubricating oil compositions are requested by receiving specification requirements. These lubricating oil compositions are formed by selecting from the database entries corresponding to compositions having specifications comparable to the received specification requirements, formulating a new lubricating oil composition to comply with received specification requirements, and testing the new lubricating oil composition for compliance with received specification requirements. The formation steps are repeated until full compliance with received specification requirements is achieved; and correlating the composition to actual engine performance.

BRIEF DESCRIPTION OF THE DRAWINGS

The foregoing and other objects, aspects, and advantages of the present invention will be better understood from the following detailed description of preferred embodiments of the invention with reference to the accompanying drawings that include the following:

Figure 1a is a block diagram of a method of formulating lubricating oil compositions of the prior art;

Figure 1b is a block diagram of a method of formulating lubricating oil compositions of the present invention;

Figure 2 is a graph of an implementation of transformed product development process showing transition to improved testing throughput of the present invention;

Figure 3 is a graph of a product development modeling framework showing the need for high product variations and high complexity and product and application environment of engine oil additives;

Figure 4 is a graph of a formulation landscape and its improvement with increase in test volume and formulation variations and the ability to make confident inferences from test results; and,

Figure 5 is a block diagram of the novel product development process for lubricating oil composition of the present inventive invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT(S)

Figures 1a and 1b illustrate a comparison of the existing technology process flow 110 from the discovery/synthesis step 112 to the core technology development step 114, through the formulation step 116 to the qualification of products step 118 (Figure 1a), with the inventive transformed technology and associated program logic control (PLC) process 120 (Figure 1b). The goal of the process 120 of the present invention is to rapidly formulate and test orders of magnitude more product variations. This will enable multiple generation products with different chemistries to be developed in parallel.

The purpose of the process 120 of the present invention is still the same as that of the current technology process, namely to create molecules and formulations that correlate with real engine performance called qualification. The inventive method substantially reduces the time to get from the discovery step 122 to the qualification step 124 because portions of research 126, 128, 130 will run in parallel. That is, chemists will synthesize and subsequently evaluate materials in parallel with formulator bench testing. In this way, much of the data generated by the chemists is directly applicable to the new specifications. This is quite different from the existing process 110, where a different set of formulations and bench test for testing newly made components is compared to those tests being run by formulators. Thus, because there is more efficiency, the time to market has contracted and the probability of success increased, component/product and aligned strategies are standardized. As an outcome, the strategic relationships with customers are strengthened.

Figure 2 illustrates a graph of implementation of the transformed product development process 208 plotted on an axis of a "test volume and formulation variations"

7

200 and an "ability to make confident inferences from test results" 202. The ability to make confident inferences from test results is a function of correlation to engine tests and statistical inferencing. Both improved testing throughput and test correlation/test inferencing must be addressed to achieve maximum results. This will require an integrated software and hardware approach. From the standpoint of cost, there is usually a preference for software solutions versus hardware solutions. It is usually less expensive to simulate with software than to develop and operate hardware solutions. However, these economics must be balanced by relative efficacy of the approaches.

To improve from a current 204 position on a graph 208 to a target 206 position, the present invention proposes performing the steps of improving a current best practice bench tests/process at step 210, apply quantitative structure activity relationship (QSAR) models and advanced informatics at step 212, perform high throughput formulation, testing and advanced informatics at step 214, and perform enhanced correlation high throughput tests at step 216.

Figure 3 illustrates a graph of product development modeling framework 300 plotted on an axis of "product variations" 302 and "relative complexity of a base product or application environment" 304. The graph 300 shows a need for high volumes of tests or high throughput experimentation (HTE) 306 increasing along the product variations axis 302, with catalysis 308 and engine oil additives 310 requiring higher volume of tests than polymers 312 and pharma products 314.

It can also be seen from the graph 300 that as the relative complexity of a base product or application environment 304 increases so do requirement for advanced statistical modeling and design of experiments (DOE) 316. In other words, the engine oil

additives 310 and pharma products 314 have a higher requirement for statistical modeling and DOE 316 than do the catalysis 308 and polymers 312.

Conversely, the ability to model molecularly from "first principles" 318 decreases in proportion to the complexity of a base product or application environment 304 and that the catalysis 308 and polymers 312 are better suited for molecular modeling from "first principles" 318 than the engine oil additives 310 and pharma products 314. Lubricating oil additives or lubricating oil compositions 310 product development transformation requires both enhanced statistical modeling/DOE approaches and HTE. Over time, integration of molecular modeling enhances the product's value even more.

Figure 4 shows a graph of "increased visibility of the formulation landscape" 400. Improving the visibility is critical to improving formulations using both current chemistry and new chemistry. Key dimensions of improvement are increasing test data points and improving inferencing abilities for a given body of test data. The graph 400 is plotted on an axis of a "test volume and formulation variations" 402 and an "ability to make confident inferences from test results" 404. The series of optimization drawings 406-412 are associated with different points along increased visibility of the formation landscape graph 400. They illustrate that it is only at a point of global optimization shown in drawing 412 that it can be ascertained that the point of current optimization, of points along the increased visibility of the formulation landscape 400 displayed by drawings 406, 408 and 410, is not the best result.

Figure 5 illustrates the inventive product development process flow 500 from conception of an idea for a lubricant. The process 500 steps include at least:

a computational work (in Silico) portion 502 comprising:

quantum mechanical (QM) molecular modeling of an idea step 504;

developing descriptors for what is important to test the QM model 506;

formulating compositions having specific characteristics 508; and

a physical portion 510 comprising:

bench testing or high throughput experimentation (HTE); and

actual engine performance.

The overall process 502 and 510 has feedback based on results and looping back 500 when things don't work as evidenced by results of the HTE and engine performance. Improving predictability of performance results at reasonable costs requires continuous improvement across in Silico modeling and physical testing processes.

Molecular Modeling

During the component QM-Level molecular modeling 504, a hypothesis exists about the possible mechanism of how a reaction mechanistically operates. For example, a tester working on a new anti-oxidant may hypothesize how peroxides form in an operating internal combustion engine and further how such peroxides are subsequently destroyed before they can do damage to a lubricating oil. To support such hypothesis the tester proposes a chemical mechanism by which such chemical process occurs. The proposed chemical mechanism is in Silico tested using methods that may be embodied in computer programs for simulating technologies. Such methods relate molecular and solid state properties to the interactions of electrons. One method from Accelrys Inc., a wholly owned subsidiary of Pharmacopeia Inc, http://www.accelrys.com/technology/qm/, works by approximating a solution to the Schrodinger equation.

The Accelrys Inc.'s QM methods predict electronic and molecular structure and energetics. Because they are based on the fundamental equation of quantum chemistry, QM methods are much more accurate than approximations such as atomistic simulation. Vibrational modes and excited states can be predicted, enabling analysis of spectral properties. The ability to predict electronic structure enables QM to study transition states and molecular orbitals - both critical to the understanding of reaction chemistry.

By iterative studies with this software the tester thus derives a plausible mechanism for the peroxide decomposition. This involves first principle calculations of, for example, transition states, bond lengths, bond angles, etc. This, for example, gives the tester a more detailed understanding of the chemical mechanism by which peroxide destruction, and therefore protection of the oil from oxidation, occurs. With such improved knowledge the tester is thereby better able to conceive of new molecules for synthesis that will improvements over the currently available antioxidants.

Deriving Descriptors

The tester is now describing descriptors for molecules that will help him build a better antioxidant. As more and more descriptors are developed the tester is building up a Quantitative Structure Activity Relationship (QSAR) library of relevant factors that are important in predicting effective from less effective antioxidants. A QSAR library or model is a multivariant mathematical relationship between a set of physicochemical properties (descriptors) and a property of the system being studied, such as the chemical reactivity, solubility, or mechanical behavior.

An example of such a program is Accelrys Inc.'s "C^2·QSAR+" software described at http://www.accelrys.com/cerius2/qsar_page.html. That and similar programs facilitate the graphical analysis of QSAR models to assist in the tester's perception of subtle chemical relationships in fields as diverse as drug discovery, polymers, and materials science. These programs integrate a wide range of regression and analysis technologies. Existing experimental data and simulation results can be used to predict the activities or properties of novel compounds thereby enabling prioritization of synthetic programs.

Formulation of Compositions Having Specific Characteristics

The tester must now synthesize sufficient numbers of molecules to validate his or her original hypothesis and test these molecules in a range of formulations to determine their actual properties. Chemists and robotics may each be used to help to create these molecules.

Bench Testing or (HTE);

The testing is done over a wide range of possible formulations with the HTE platform. It is necessary to study these components over a range of formulations because individual lubricating oil additives may interact with each other to either enhance or reduce their predicted activity.

At any step of this process the tester might need to loop back and restart the process of deriving descriptors 506 or all the way to the re-start the QM-Level molecular modeling. Eventually, the tester will collect a massive amount of data from the HTE. This data is then used to build a library of information on components in formulations. The library is quite extensive, as it comprises each component in combination with

numerous other components and the entire mixture in multiple base oils. In particular,

the library provides separate results for each formulation versus individual bench tests.

The management of the data can be achieved by a high-throughput software using a

relational database like Oracle or Sybase. An example of such software is the

"CombiMat" program from Accelrys Inc., described on

http://www.accelrys.com/mstudio/ms_matinformatics/combimat.html.

From the standpoint of cost, there is usually a preference for software solutions

versus hardware solutions. It is typically less expensive to simulate with software than to

develop and operate hardware solutions. However, these economics must be balanced by

relative efficacy of the approaches. Therefore, everything else being equal, there would

be a preference to employ:

1. Molecular modeling that predicts component and product performance from,

"first principles". This element includes software only.

2. Comprehensive statistical inferencing that can correlate performance across a

chain of tests, e.g., basic lab tests, bench tests, and engine tests, with the structural

characteristics of the molecules of the product. A general approach to this inferencing is

called Quantitative Structure-Activity Relationship (QSAR) modeling. This element

includes software only.

3. High Throughput Experimentation (HTE) approaches, which enables vastly

more physical tests on components and products than through conventional means. This

test, includes hardware with associated informatics software.

Actual Engine Performance

Advanced statistical correlations of bench test data versus actual engine test data can now be performed. This is achieved through software like the Formulation Assisting Software Toolkit from Accelrys Inc. described at http://www.accelrys.com/mstudio/ms_matinformatics/fast.html. Such software is a workflow solution for design of formulated products that streamlines the data, information, and knowledge flow that is necessary to successfully formulate products. It is likely that no single bench test will correlate perfectly with the engine test so linear or higher order regressions of multiple wear and oxidation tests will be used to produce a weighted combination that best correlates with the engine test results. Such regression analysis is used to make predictions about a single value. Linear or higher order regression involves discovering the equation for a line or curve that best fits the given data. Such equation is then used to predict engine test results for additional formulations.

This then allows the tester to optimize the final formulation for cost/performance. That is the performance is moved in a direction that is predicted to still pass the performance but at the lowest possible cost.

As another example, the QM computations drive the in Silico modeling process that provides insight into the mechanism of problems. For example, if one wishes to prepare better antioxidants it may be important to understand the mechanism by which antioxidants function. In studying the mechanism by which certain transition metal function as a peroxide inhibitor, a quantum mechanical study of the transition states for decomposition of such peroxides may be done evaluated in Silico.

As the mechanism is further understood, so are the molecular descriptors that are useful for chemists to make, e.g., improved antioxidants. These molecular descriptors are then utilized in the QSAR software. As one further understands the QM, a larger set of descriptors is built up for helping to predict a family of chemical structures that will further improve oxidation performance. At this point the work is still performed at a computational or an in Silico level.

It may be found, for example., that one antioxidant does not function as well alone as it does in combination with other antioxidants. Further, it may be possible to regenerate certain antioxidants thereby creating a catalytic cycle and thus more efficiently utilize the combination of antioxidants compared to either one separately. By this in Silico method it becomes possible to construct, a new formulation for meeting specific performance criteria. Throughout this process new chemicals are being developed which test pre-conceived models and provide a feedback loop to validate prior QM models and thereby further improve the QSAR descriptors.

With a sufficient base of knowledge it becomes beneficial to run HTE to validate such in Silico created models using bench testing. To analyze this data, it becomes useful to have advanced statistical and informatics software so trends and correlations of data to performance can be evaluated. As more testing data is understood the feedback continues to the QM and QSAR, so that both become improved as real data validates or refutes earlier theories.

Finally, with high volume and improvements in bench test results one moves to the engine test performance correlation.

While the invention has been shown and described with reference to certain preferred embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the spirit and scope of the invention as defined by the appended claims.